Methods

Mice, superovulation, timed breeding and staging

B6C3F1/J 3-week old female mice were purchased from Jackson Laboratories (Stock no: 100010). Females were injected with 5 IU pregnant mares' serum gonadotrophin (Sigma) and 48 hours later with 5 IU human chorionic gonadotrophin (Sigma). Females were paired overnight with *Runx1*$^{IRES\text{-}GFP}$ (*Runx1*^{tm4Dow}) [1] homozygous males. Embryos were staged by counting somites. Embryos displaying abnormal development were discarded.

Isolation of embryonic tissues

Yolk sacs (YS) were removed from E9.5 and E10.5 embryos, and vitelline vessels kept with the embryonic portion of conceptuses. The head, cardiac and pulmonary regions, liver, digestive tube, tail and limb buds were removed. The remaining portion containing the dorsal aorta in the aorta-gonad-mesonephros region, portions of somite, umbilical and vitelline vessels were collected as "arteries". The dissected tissues were digested for 1hr in 0.125% collagenase (Sigma) in PBS/10% FBS (Gibco)/1% Pen/Strep, washed, and resuspended in PBS/10%FBS/PS.

Flow cytometry and RNA preparation

A total of 651 embryos and yolk sacs were collected for E9.5 tissues and 264 embryos and yolk sacs for E10.5 tissues for RNA-Seq. Cells were labeled with the following antibodies from eBiosciences: CD41-PE (AB_2538354), CD31-PE-cy7 (AB_469616), CD45-APC (AB_469392), and Kit-APC-efluor780 (AB_1272177). DAPI (Invitrogen) was used for viability. Samples were sorted on a BD Influx into Trizol-LS (Thermo) or PBS/20%FBS/PS/25mM HEPES. RNA was isolated using Qiagen RNeasy Microkit with DNAse step performed on column or using Turbo DNA-*free*^{TM} Kit (Thermo).

Endothelial and hematopoietic validation assays

Endothelial tube assays were performed as described by Medvinsky et al. [2]. Sorted populations were cultured on OP9 stromal cells on a 0.1% gelatin coated 96-well plate (8500 OP9 cells/well) in aMEM, 10% FBS (Gibco 16000-044) and 50 ng/mL VEGF (Peprotech) for 4 days. Immunohistochemistry was performed using 5μg/mL anti-mouse CD31 (BD) and visualized using Vectastain ABC-AP kit and Vector Blue Alkaline Phosphatase Substrate Kit (Vector Labs). $CD31^+$ endothelial tubes were counted on a light microscope. Hemogenic potential was assessed by culturing sorted cells in limiting range on OP9 stromal cells supplemented with 10% FBS (Gibco) and 10 ng/mL each SCF, IL-3, Flt3L and IL-7 (Peprotech) for 9-11 days as described by Swiers et al. [3]. Each well was assayed for the presence of $CD45^+$ cells on a BD LSRII flow cytometer. The frequency of HE events was calculated using ELDA software [4]. Hematopoietic colony assays were performed in M3434 (StemCell Technologies) and counted after 7 days.

Transcriptome assembly and expression level estimate from read counts

Paired-end sequencing reads were mapped to the reference mouse genome (release mm9) using Tophat2 [5]. Only uniquely mapped reads with fewer than 2 mismatches were used for downstream analyses. Transcripts were assembled using Cufflinks [6] using mapped fragments outputted by Tophat. Ensemble (release 66) was used as the source of annotated genes and transcripts. Normalized transcript abundance was computed using Cufflinks and expressed as FPKM (Fragments Per Kilobase of transcripts per Million mapped reads). Gene-level FPKM values were computed by summing up FPKM values of their corresponding transcripts [6]. Following previous studies [7], we used a FPKM value of one as the cutoff for expressed genes, which represents approximately one copy of RNA per cell. RNA-Seq data

reproducibility was assessed by computing Spearman correlation of gene expression between a pair of biological replicates. Genes with zero read counts in all biological replicates were excluded from the correlation calculation.

Identification of differentially expressed genes and pathway enrichment analysis

FeatureCounts [8] was used to summarize read counts for each gene. With normalized read counts for each gene, EBSeq [9] was used to detect significantly differentially expressed genes with a false discovery rate (FDR) cutoff of 0.05 and a fold change cutoff of 1.5. We performed three pairwise comparisons: artery hemogenic endothelium (HE) vs. artery non-hemogenic endothelium (E), yolk sac (YS) HE vs. YS E, and artery HE vs. YS HE. Enriched signaling pathways were identified using Enrichr for each comparison using the differentially expressed genes [10].

Construction of condition-specific transcriptional regulatory networks (TRNs) using RNA-Seq data

A recent study assessed the performance of 35 computational methods for inferring TRNs using gene expression data alone [11]. The study revealed that no single inference method performs optimally across all data sets. In contrast, integration of predictions from multiple inference methods shows robust and high performance across diverse data sets. We thus sought to build a consensus TRN by using five top performing inference methods, including a method based on Pearson correlation, the context likelihood of relatedness (CLR) method [12], Inferelator [13], trustful inference of gene regulation using stability selection (TIGRESS) [14], and gene network inference with ensemble of trees (GENIE3) [15].

We collected a set of 146 gene expression profiling data for cells of the hematopoietic

system from the previously published five studies [16-20], and the current study. Before combining datasets from different studies, we performed quantile normalization using the RMA algorithm [21] and removed batch effects using the ComBat algorithm [22]

To infer a conditional-specific TRN using the consensus approach, two consensus TRNs are first built, G_{all} based on using all samples in the expression compendium, and $G_{all\text{-}condition}$ based on using all samples minus the samples of interest. To obtain the condition-specific TRN $G_{condition}$, all edges in $G_{all\text{-}condition}$ ($E_{all\text{-}condition}$) are eliminated from G_{all} (Fig. 2D).

Prioritization of key transcription factors in a TRN

We used the constructed TRNs to identify key transcription factors (TFs) regulating a developmental transition. To this end, we assumed that a key TF tends to have a larger impact on the full set of differentially expressed genes during the transition. From a network perspective, this means the regulatory role of a key TF tends to be propagated to a larger set of differentially expressed genes in the TRN, either via direct or indirect connections. Based on this assumption, we computed a distance between two genes, $\boldsymbol{i}$ and $\boldsymbol{j}$, in the TRN as following: $\boldsymbol{W_{i,j} = 1 - \frac{log_{10}(p_i)+log_{10}(p_j)}{2*log_{10}(p_{min})}}$, where $\boldsymbol{p_i}$ and $\boldsymbol{p_j}$ are the differential expression p-values for gene $\boldsymbol{i}$ and $\boldsymbol{j}$, respectively. $\boldsymbol{p_{min}}$ is the minimum differential expression p-value among all genes in the TRN. With the distance-weighted TRN, we calculate an average shortest distance between a given TF and all differentially expressed genes in the network. We computed the pairwise shortest path using the Dijkstra's algorithm [23]. Statistical significance of average shortest distance is computed using a null distribution computed on randomized networks. To obtain randomized networks, the edge weight of the real network was shuffled but with node degree preserved. This procedure was repeated 1,000 times to generate a set of randomized networks.

References

[1] Lorsbach RB, Moore J, Ang SO, Sun W, Lenny N, Downing JR. Role of Runx1 in adult hematopoiesis: analysis of Runx1-IRES-GFP knock-in mice reveals differential lineage expression. Blood. 2004;103:2522-2529.

[2] Medvinsky A, Taoudi S, Mendes S, Dzierzak E. Analysis and manipulation of hematopoietic progenitor and stem cells from murine embryonic tissues. Curr Protoc Stem Cell Biol. 2008;Chapter 2:Unit 2A 6.

[3] Swiers G, Baumann C, O'Rourke J, et al. Early dynamic fate changes in haemogenic endothelium characterized at the single-cell level. Nature communications. 2013;4:2924.

[4] Hu Y, Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. Journal of immunological methods. 2009;347:70-78.

[5] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.

[6] Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511-515.

[7] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5:621-628.

[8] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923-930.

[9] Leng N, Dawson JA, Thomson JA, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics. 2013;29:1035-1043.

[10] Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44:W90-97.

[11] Marbach D, Costello JC, Kuffner R, et al. Wisdom of crowds for robust gene network inference. Nat Methods. 2012;9:796-804.

[12] Faith JJ, Hayete B, Thaden JT, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. 2007;5:e8.

[13] Bonneau R, Reiss DJ, Shannon P, et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol. 2006;7:R36.

[14] Haury AC, Mordelet F, Vera-Licona P, Vert JP. TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. BMC Syst Biol. 2012;6:145.

[15] Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS One. 2010;5.

[16] Li Y, Esain V, Teng L, et al. Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production. Genes Dev. 2014;28:2597-2612.

[17] Bagger FO, Rapin N, Theilgaard-Monch K, et al. HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. Nucleic Acids Res. 2013;41:D1034-1039.

[18] Chambers SM, Boles NC, Lin KY, et al. Hematopoietic fingerprints: an expression database of stem cells and their progeny. Cell Stem Cell. 2007;1:578-591.

[19] McKinney-Freeman S, Cahan P, Li H, et al. The transcriptional landscape of hematopoietic stem cell ontogeny. Cell Stem Cell. 2012;11:701-714.

[20] Di Tullio A, Vu Manh TP, Schubert A, Castellano G, Mansson R, Graf T. CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B

cells into macrophages involves no overt retrodifferentiation. Proc Natl Acad Sci U S A. 2011;108:17016-17021.
[21] Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4:249-264.
[22] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118-127.
[23] Dijkstra EW. A note on two problems in connexion with graphs. Numerische Mathematik. 1959;1.